UNITED
NATIONS

E

Economic and Social Council

Distr.
LIMITED

E/CONF.83/L.35
21 January 1991

ENGLISH ONLY

TWELFTH UNITED NATIONS REGIONAL
CARTOGRAPHIC CONFERENCE FOR
ASIA AND THE PACIFIC
Bangkok, 20-28 February 1991
Item 7 (b) of the provisional agenda*

CARTOGRAPHIC DATA RETRIEVAL, ANALYSIS, DEPICTION, PRESENTATION AND PRODUCT GENERATION: DIGITAL DATA BASES

The United States Geological Survey-United States Bureau of the Census cooperative digital mapping project: A unique success story

Paper submitted by the United States of America**

---

* E/CONF.83/1.

** Prepared by Bruce Y. McKenzie, United States Geological Survey and Robert A. LaMacchia, United States Bureau of the Census.

Between December 1983 and June 1987, the U.S. Geological Survey (USGS) and the U.S. Bureau of the Census (Census Bureau) cooperated to establish a 1:100,000-scale transportation and hydrography digital data base. The magnitude of the task of converting to digital data the transportation and hydrography information contained on 1,823 1:100,000-scale quadrangles and the short timeframe for accomplishing this project required the development and implementation of state-of-the-art digital production systems in both agencies. The completed 1:100,000-scale digital data files are now part of the National Digital Cartographic Data Base (NDCDB) maintained by the USGS. The USGS distributes these files as U.S. GeoData 1:100,000-scale Digital Line Graphs (DLG's). The transportation and hydrography data are being distributed in 30- by 30-minute blocks. These 30- by 30-minute areas correspond to the east or west halves of the 1:100,000-scale maps. These data are topologically structured, attribute coded and available in standard and optional DLG format. Because 1:100,000-scale DLG's have a topologically structured format in which relationships among features are encoded, the data can be used for general spatial applications, to perform statistical calculations, or for tabulations.

The Census Bureau is using the 1:100,000-scale transportation and hydrography digital data as the cartographic base for their Topologically Integrated Geographic Encoding and Referencing (TIGER) System, which covers the Nation, Puerto Rico, Virgin Islands, and outlying areas of the Pacific. The Census Bureau has added information such as feature names and political and Census statistical boundaries and codes to the TIGER files for use in the 1990 decennial and subsequent economic censuses and surveys. For the 1990 decennial census alone, the Census Bureau will produce over 2 million individual map sheets from these data.

The joint venture between the USGS and the Census Bureau has great economic significance. The 1:100,000-scale digital cartographic data base combines the qualities of sufficient detail, adequate content, and nationwide coverage that will allow and encourage increased use of spatial analysis methods. For example, after 1990, the data reported by the Census Bureau from the 1990 decennial census can be related to features included in the 1;100,000-scale data base. The USGS and the Census Bureau are making all of these data available through both agencies' information outlets. This compatibility of census data with the USGS digital cartographic data base can only be of increased benefit to all Federal, State, local, and private users.

THE U.S. GEOLOGICAL SURVEY-U.S. BUREAU OF THE CENSUS COOPERATIVE DIGITAL MAPPING PROJECT: A UNIQUE SUCCESS STORY

THE BEGINNING

Initial Test Project

The need for a digital cartographic base as the "heart" of the Census Bureau's full range of cartographic and geographic support processes necessary to serve the data collection, tabulation, and dissemination needs of the 1990 decennial census, led the Census Bureau to begin discussions with the USGS. These discussions led to the formation of an interagency task force and initiation of test projects to determine the feasibility of building the required digital data base.

With the formation of the interagency task force in 1981, the USGS and the Census Bureau wanted to answer the question of comparability of between a digital cartographic file based upon digitizing the 1:24,000-scale map sheets and a digital cartographic file based upon the 1:100,000 map sheet. Two 1:100,000-scale map sheets were selected for the test: the USGS raster-scanned the map sheets and created vector files, and the Census Bureau manually digitized the corresponding 1:24,000-scale map sheets using it's Intergraph equipment. The USGS provided the 1:100,000-scale vector file to the Census Bureau which then plotted the data from the two separate files in different colors at the 1:24,000-scale. This test showed that the 1:100,000-scale data compared favorably at the 1:24,000-scale and could be further enlarged and used to produce maps for enumerator use.

This test only answered the first question--that data digitally captured from the 1:100,000-scale maps could be used to produce an acceptable large-scale map. The question which remained to be answered was whether the USGS and the Census Bureau could undertake this process for the entire Nation. The two agencies decided to undertake a pilot project to answer this second question.

Florida Pilot Project

In early 1983, a cooperative digital pilot project was begun to collect and process the transportation and hydrographic data shown on the 48 1:100,000 scale maps that cover the State of Florida. The purpose of the project was to enable the USGS to develop and test new production procedures and software, to incorporate the Scitex Response scanning and editing system into its current digital production system, and to enable the Census Bureau to develop the necessary software and production, procedures to encode and structure data for incorporation into the National Digital Cartographic Data Base. This pilot project was designed with (1) the USGS performing the initial collection and processing of all transportation and hydrographic information using the Scitex system, (2) the Census Bureau performing all attribute coding and structure of the roads data, (3) the USGS performing all the attribute coding and structuring of the hydrographic and other transportation data, and (4) the USGS performing final data verification, testing, and

/...

storing, of the information. Each phase of this project provided valuable insight on data gathering and processing procedures. It also provided data sets for evaluation and testing by both agencies.

The Official Start

The results of this pilot project led to the commitment by both agencies, through an agreement signed in December 1983, for completion of a conterminous U.S. 1:100,000-scale transportation and hydrography digital data base by mid-1987. To meet the deadline, both agencies proceeded with the implementation of complementary 1:100,000-scale digital production systems. The data digitized to build this data base were all the transportation (roads, railroads, powerlines, pipelines, and airports) and hydrographic (streams, rivers, water bodies) features shown on the 1,823 1:100,000-scale quadrangles covering the conterminous United States.

PRODUCTION SYSTEMS--AGENCY RESPONSIBILITIES

The production procedures and software that were developed as a result of the cooperative pilot project for the State of Florida have been described in considerable detail in other papers. The following paragraphs summarize the production process during the project, which improved the quality of the data and resulted in efficiencies in the data collection procedures.

Base Map Operations

The fundamental graphic product required for digital data collection was the USGS 1:100,000-scale map series. The framework and content of the 1:100,000-scale base map is derived primarily from 1:24,000-scale maps, with updates during the production process. This production process generally has the following phases: (1) reducing larger scale maps to 1:100,000-scale and mosaicking on the Universal Transverse Mercator projection, (2) updating the mosaicked base, (3) scribing the planimetric feature separate manuscripts, (4) scribing the contour manuscript, and (5) printing the complete topographic edition.

Since the 1:100,000-scale mapping program began in 1976, the map series has been predominately topographic contour editions. When the cooperative digital project began in December 1983, only 960 of the required 1,823 maps were available for digitizing. In order to ensure that the remaining base maps were available for digitizing, emphasis was shifted from preparing topographic editions to producing planimetric bases without contours. This compilation change coupled with increased personnel resources enabled conterminous U.S. coverage to be achieved in October 1986; 5 years ahead of the original mapping goal.

Digital Operations

As mentioned previously, in order to meet the mid-1987 deadline, complementary high-volume digital production systems were established in each agency. The USGS had

dedicated to this project effort, 2 Scitex scanners; 15 Scitex R-280 editing stations; 26 Intergraph Interactive editing/tagging stations; 13 Altek digitizing tables; and 4 Gould 9780 minicomputers. This equipment was dispersed among four regional mapping centers: Eastern Mapping Center in Reston, Va., Mid-Continent Mapping Center in Rolla, Mo.; Rocky Mountain Mapping Center in Denver, Colo.; and Western Mapping Center in Menlo Park, Calif. The Census Bureau employed the use of a Sperry 1100/74 and subsequently a Sperry 1100/92 mainframe computer. Connected to these computers were 18 Tektronix 4115 computer graphic terminals (later upgraded to Tektronix 4125 terminals) located at the Census Bureau in Suitland, Md., and an additional 8 Tektronix 4115/4125 terminals located in the Census Bureau's Boston Field Digitizing Site.

Each agency staffed most of this equipment in three 8-hour shift operations per day. The USGS had the responsibility for digitizing all transportation and hydrographic features and assigning feature classification codes to all data except roads; the Census Bureau had the responsibility for assigning feature classification codes to the road data.

Two basic production procedures were used by USGS to capture the data: manual digitizing and automated scanning techniques. Due to lack of density, the miscellaneous transportation data (pipelines, powerlines, airports, etc.) were digitized using manual line-following techniques, subjected to both manual and computer editing, processed through topologic structuring software, and archived. The hydrography and road files were typically much more dense and complex. They were, therefore, scanned by machine; edited via both batch and interactive procedures; and through software, converted from raster to vector data. At this point the hydrography and roads categories took two separate processing paths. The hydrography data were tagged with appropriate attribute codes and further edited using interactive graphic systems by the USGS. The road data was sent to Census for attribute tagging geometric corrections, and then returned to USGS for review and acceptance.

Following the completion of the editing and tagging operation, both data sets were processed through the Unified Cartographic Line Graph Encoding System (UCLGES) or more recently the DLG Production System (PROSYS) software resident in each production center. These systems perform all final data structuring and verification of DLG files. Both systems also perform several logical checks to verify topology and the proper use of attribute codes. Errors were corrected either through a batch editing routines or the data sets were returned to the producer (USGS or Census Bureau) for correction. When a file was completely verified, the files were archived and prepared for delivery to the Census Bureau.

At this juncture, both categories for each 30- by 60-minute quadrangle were delivered on tape to the Census Bureau where a final check was conducted to ensure that both categories were internally consistent. This check was accomplished using software that first horizontally integrated the individual 7.5-minute quadrangle data files that comprised each 30- by 60-minute quadrangle, and then vertically integrated the four resulting data files to create a single topological file in a preliminary TIGER structure. After successfully passing this final check, the 1:100,000-scale digital data were archived by USGS for sale to the public through the National Digital Cartographic Data Base (NDCDB).

Problems Related to File Production

The project was not without problems. Just ironing-out the logistics associated with the transfer of data and materials between the two agencies, and even within each agency, created some complex and sensitive situations. In addition to the general problems associated with expediting the production of the 1:100,000-scale graphics, there were numerous technical problems that arose in establishing a digitizing, editing, and processing routine to convert these graphics into digital form. These issues can be grouped into two primary areas: (1) data problems and (2) quality-control procedures. The area of data problems includes four categories:

1) System implementation - problems introduced due to techniques and software development during the initial stage of the project, and primarily involved developing and adapting software for utilization on different versions of the same equipment.

2) Data Processing - problems introduced due to clerical errors and miskeying of input data.

3) Interpretation - problems introduced due to operator misunderstanding of attribute coding standards or map symbology.

4) Geometric - problems introduced by system digitizing and processing routines and resulted in the automatic generation of extraneous lines and gaps in line segments.

The first three categories were short-term problems that eventually disappeared with the increasing experience of personnel and stabilization of the production process. The fourth category, geometric problems, was the area where the most effort was expended for resolution. New software had to be developed and procedural changes in both agencies were required. In some instances, completed data sets had to be reprocessed and reverified.

Since 1:100,000-scale digital data comprised a new product, the establishment of quality-control procedures was essential to insure that the final product met not only the needs of the Census Bureau but also met the Geological Survey standards for entry in the NDCDB. Several procedures were adopted to minimize the potential for occurrence of errors in data. The most significant was the implementation of a criteria for verification and acceptance of corrected and tagged road data. Quality-control procedures included a recheck for all attribute codes for randomly selected sections of a 1:100,000-scale quad; the use of topology-verification software routines; visual inspection of plots to check the digital data against the graphic; and a final delivery acceptance/check of each data set (hydrography and transportation) performed by the Census Bureau through a process called horizontal and vertical integration of the data.

CENSUS BUREAU USE OF THE DIGITAL DATA

The Census Bureau is using the digital data from this cooperative project, as the "heart" of a bold new system to modernize its cartographic and geographic processes. This new data base, called the TIGER System, will support the Census Bureau's future demographic and economic data collection, processing, and tabulation activities, beginning with the 1990 decennial census. To use the 1:100,000-scale digital data for this purpose, the Census Bureau enhanced and reformatted the USGS DLG data to support its field and processing operations. This involved completing seven major tasks:

1. The Census Bureau's Geography Division vertically integrated the four separate data categories for each 1:100,000-scale map in an automated process that determines where the hydrographic, railroad, and miscellaneous transportation features intersect the road features and each other. This process results in a single topological data base where a Census Bureau feature class code identifies each line and a numeric code identifies each area enclosed by a set of lines.

2. While the process of building the digital file was underway, the Census Bureau's 12 regional office geographic staffs compiled map update information on 1:24,000-scale USGS map sheets using sources gathered from local government officials. These new features included roads, railroads, and hydrographic features. Appended to each feature, both old and new, was a key number that represented the name of the feature where the geographic staff could determine the name. Where multiple names were known, one name was designated as the primary name and all others were designated as alternates.

3. The Census Bureau created individual 7.5-minute files from the topological data base for use on their standalone microcomputers in four Field Digitizing Sites (FDS) in Boston, Atlanta, Dallas, and Denver. The data were downloaded from the Sperry mainframe computer to 8-inch floppy disk's for use in the Tektronix 4125 microcomputer graphic workstations. The FDS staff updated the digital file based upon the information posted to the 1:24,000-scale maps; this update included the new features and feature names. A total of 60 Tektronix 4125 microcomputer graphic workstations were used for this work. The updated files were returned on 8-inch floppy disk's to the Geography Division where the files were uploaded to the Sperry for further processing.

4. Staff in the Census Bureau's Data Preparation Division in Jeffersonville, Ind., annotated maps, primarily using USGS quadrangles, showing the 1980 census geographic area boundaries. The geographic area boundaries and codes shown on these maps were entered by Census Bureau staff, using an additional 20 Tektronix 4125 microcomputer graphic workstations, in Jeffersonville, Suitland, and the four Field Digitizing Sites into the growing preliminary TIGER data base. Again, the file transfer between these sites and the Sperry computer was by 8-inch floppy disk.

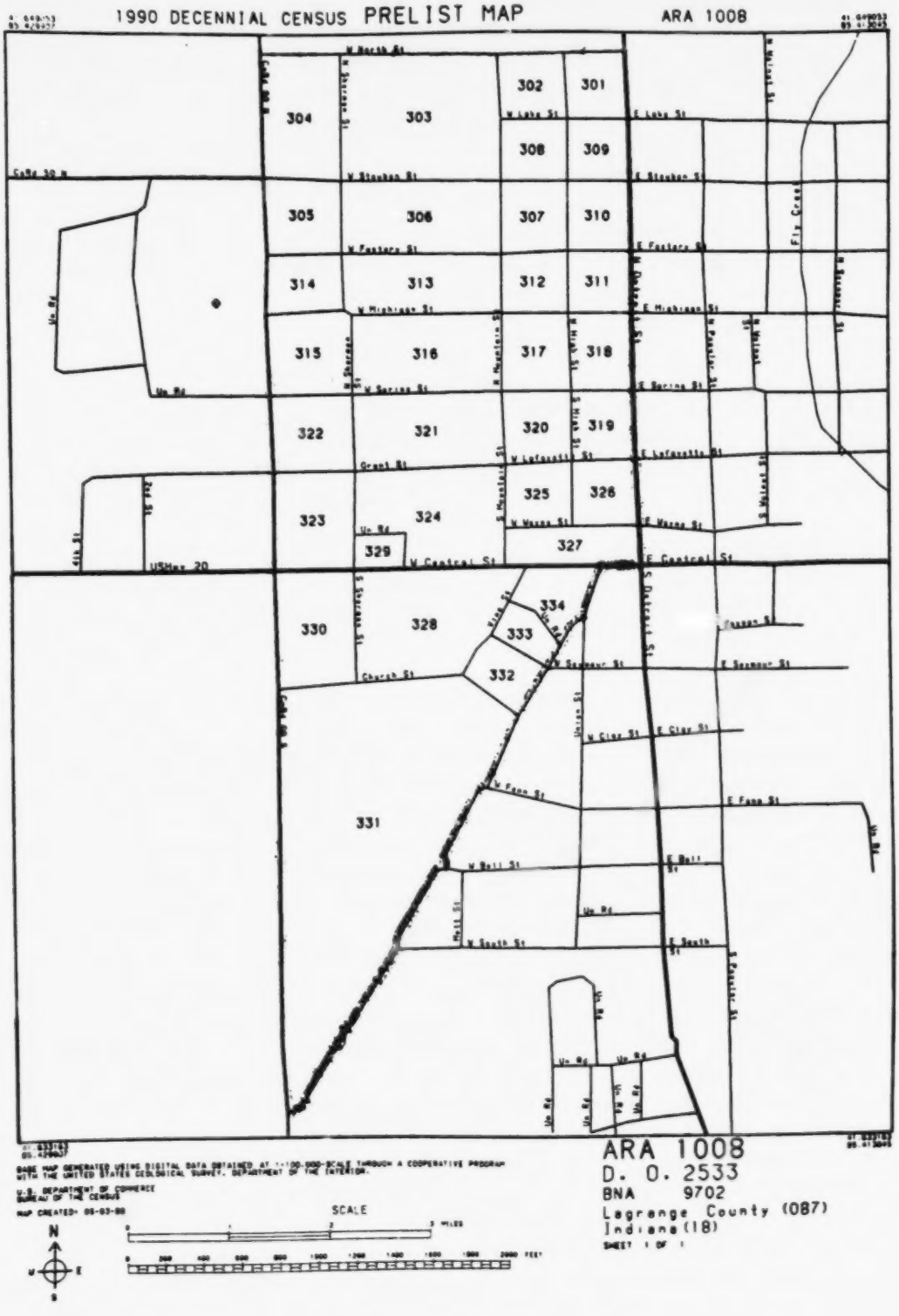
/...

5. Once the 1980 census geographic area codes were in the preliminary TIGER data base, the separate 7.5-minute files were horizontally integrated. During this step, all of the 7.5-minute "cells" that comprise a county are put together into a single file and all lines that cross a cell boundary are edge matched. Once the horizontal integration is complete, the county file is created by "trimming" away all of the polygons that are outside of the county. The outer edge of this file is now fixed for the duration of the census to ensure that adjacent files exactly match.

6. Using copies of maps produced from the county files, the geographic staff in the Census Bureau's 12 regional offices annotated the 1990 census "collection" geographic area boundaries delineated by local statistical area committees and State agencies. Staff in the Geography Division and the Boston and Denver FDS's added these additional geographic area codes to the data base for use in collecting 1990 census data. These geographic areas include census tracts and block numbering areas, block groups, and American Indian reservations. The computer assigns the 1990 census block numbers to all blocks in the country based upon an algorithm that uses the feature class code of the bounding line segments and the size of the polygons.

7. Finally, to tabulate and present the results of the 1990 decennial census, the Census Bureau added the 1990 geographic area boundaries and identification codes for statistical areas such as Census Designated Places, and Census County Divisions. The Census Bureau also corrected the political boundaries in the TIGER file to represent the January 1, 1990 Boundary and Annexation Survey (BAS). The Census Bureau conducts the BAS annually, and beginning with the 1988 BAS, sent local governments a map of their governmental unit plotted from the TIGER file. The local official used this in reviewing their boundaries and for showing the new boundaries caused by annexations or detachments.

In October 1987, the Census Bureau began the production of over 250,000 individual map sheets for the approximately 82,000 enumeration areas (block groups) from which the Census Bureau field staff created address lists for use by enumerators performing preliminary field work during the summers of 1988 and 1989. These maps were plotted on electrostatic plotters. (A sample section from an enumerator assignment map produced from the TIGER File appears as figure 1.) The Census Bureau prepared these map products using a totally automated production system that determines the scale of each map based upon the features to be displayed for the particular area. This mapping software also automatically determines when to produce "inset" sheets for areas of congested street pattern; places all of the text (street, river, and railroad names); and places all of the geographic area codes such as the block numbers in the appropriate polygons. When necessary, the mapping software determines if a feature name needs to be displayed more than once, if any census block needs "fishhooks" (a line connecting two blocks indicating that the adjacent block has the same block number), displays landmark symbols, and generates related marginal text, including map identification numbers, a map sheet index (if required), State and county names and codes, and a bar code representing the map identification information in machine-readable form.


1990 DECENNIAL CENSUS PRELIST MAP
ARA 1008
W Berrick St
E Lake St
W Lake St
E Stanton St
W Stanton St
E Fostoria St
W Fostoria St
E Michigan St
W Michigan St
E Perry St
W Perry St
E Barnes St
W Barnes St
E LaFayette St
W LaFayette St
E Wayne St
W Wayne St
E Central St
W Central St
Grant St
Church St
Union St
E Clay St
W Clay St
E Pane St
W Pane St
E Bell St
W Bell St
E South St
W South St
Mill St
S Detroit St
S Public St
S Wayne St
S Main St
S Meyer St
S Spring St
S Forrest St
Fly Creek
US Hwy 20
Un Rd
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
ARA 1008
D. O. 2533
BNA 9702
Lagrange County (087)
Indiana (18)
SHEET 1 OF 1
SCALE
N
W
E
S

/...

The Census Bureau estimates that before the 1990 census is over, it will produce over 2 million individual map sheets on their electrostatic plotters. All of these maps will be plotted without manual intervention. The only human involvement is the submittal of the computer processing runs; the mounting of tapes on the computers and electrostatic plotters, and cutting the map sheets as they come from the plotter.

AVAILABILITY AND DISTRIBUTION OF DATA

On August 4, 1987, the two agencies announced plans for the cooperative promotion of their products resulting from this project at the annual meeting of the Urban and Regional Information Systems Association. These products, known as US GeoData (for DLG's) and TIGER data respectively, can be ordered from either agency. Both agencies have information regarding the complete set of products available from this effort.

The Census Bureau completed release, in April 1989, of the first public product from the TIGER data base--the TIGER/Line file. This product is being made available in four versions: Prototype (available April 1989), Precensus (available February 1990), with Initial Voting District Codes (available October 1990), and the 1990 Census (available beginning in January 1991 with the last files released by the end of March 1991). The price for these files is based on the number of counties purchased at one time within a single state; the price varies with the version of the TIGER/Line file. The Precensus TIGER/Line file is $200 for the first county in a state and $25 for each additional county in the state ordered at the same time.

The Census Bureau expects to release additional extracts from the TIGER file in 1991 after the completion of the census. These additional extracts will include the TIGER/Data Base, an extract file that will contain the most complete set of releasable geographic and cartographic data possible in standard interchange format; TIGER/Comparability, an extract file that will contain records that equate 1980 census tracts to 1990 census tracts; TIGER/Boundary, extract files that contain coordinate data for several specific boundary sets such as county and census tract boundaries; and TIGER/GICS, a file providing the name and some related status codes for each geographic area for which the Census Bureau tabulated data in the 1990 Census.

On July 15, 1986, the USGS began distribution of US Geodata Digital Line Graphs from 1:100,000-scale maps. All the transportation and hydrography digital data collected from 1,823 1:100,000-scale quadrangles are available through the USGS NDCDB. The 1:100,000-scale DLG's are topologically structured, attribute coded, and available in standard DLG format (has an internal file coordinate system) or optional DLG format (has already been converted to Universal Transverse Mercator coordinates.) Data on transportation and hydrography are presently distributed in two packaged versions. In the first (preliminary) version, the 7.5-minute files (except for railroad, pipelines, and power lines, which were digitized and are distributed as 30-minute files) are not edge-matched but are made available for each 30- by 30-minute area of coverage (east or west halves of the 1:100,000-scale maps). For each 30-minute block, the two categories of data--hydrography and transportation--are sold separately. All hydrography (flowing water, standing water, wetlands) data is contained in one file; transportation data is contained

in three separate files (1) roads and trails, (2) railroads, and (3) pipelines, transmission lines, and airports. For the second (final) distribution version, the aforementioned 7.5-minute files have been processed through paneling, partitioning and edge-matching software and are produced and distributed as four 15- by 15-minute files for each 30- by 30-minute block, except in high-density areas, where the 15-minute file size precludes the joining the 7.5-minute files. The process of repackaging results also in a 30- by 30-minute unit of railroads or miscellaneous transportation data being partitioned into four 15- by 15-minute data files. As data sets are paneled and partitioned into 15-minute blocks, they are replacing those 1:100,000-scale 7.5-minute files currently in the NDCDB. This paneling and partitioning process is complete for 60% of the 1:100,000-scale files. The price for each 30- by 30-minute area of coverage is dependant upon the total number of 30-minute block units purchased. If 6 or more units are being purchased, the cost is a base charge of $90 plus $7 for each additional sales unit.

APPLICATIONS/FUTURE DIRECTIONS

The USGS and the Census Bureau serve rather different markets. The Census Bureau serves customers interested in portraying political and statistical areas, producing thematic maps based on 1990 census information, performing density calculations, and undertaking other related statistical data analysis. The USGS serves customers needing base map digital data and those customers who perform automated integration of spatial data with natural resource data bases in geographic information systems (GIS).

In the area of GIS's, both agencies are paying attention to the future direction and the future challenge of the joint effort. The USGS is making considerable effort to explore innovative GIS applications of spatial data bases, and also is continuing to expand the contents of the 1:100,00-scale digital data base by adding Public Land Survey System, boundaries, and hypsography digital data. The 1:100,000-scale base categories of digital data are especially well suited for regionally based applications such as mineral resource analyses, hydrologic analysis, and regional planning, within a GIS environment. The addition of Census Bureau information in the 1990's will allow spatial analysis of vast amounts of socioeconomic data and may serve as a catalyst for the widespread use of GIS technology in the United States.

CONCLUSION

The U.S. Geological Survey and the Census Bureau reached a major milestone with the completion of their cooperative project to develop a 1:100,000-scale transportation and hydrography digital data base for the United States. Reaching this milestone provides the map user community with a significant new product: 1:100,000-scale digital data for the roads, hydrography, railroads, and miscellaneous transportation features shown on the printed 1:100,000-scale maps. The nationwide digital data coverage at this scale combines the qualities of sufficient detail, and adequate content, that can serve as a framework for use in geographic information systems. The compatibility of census data with USGS digital cartographic data can only offer increased benefit to all federal, State, local, and private users.